An Audible Ultrasound Electrolarynx

Dissertation Defense

March 26th, 2014

Thank you for coming today. I have prepared a presentation covering my original work on the development of an audible ultrasound electrolarynx.

I'll briefly cover how speech is generated and what happens in a laryngectomy.

And review some speech restoration technologies focusing on the electrolarynx.

I'll then discuss active research.

Finally I'll cover the original work I've done to design and create the device and present results from our simulators and working device.

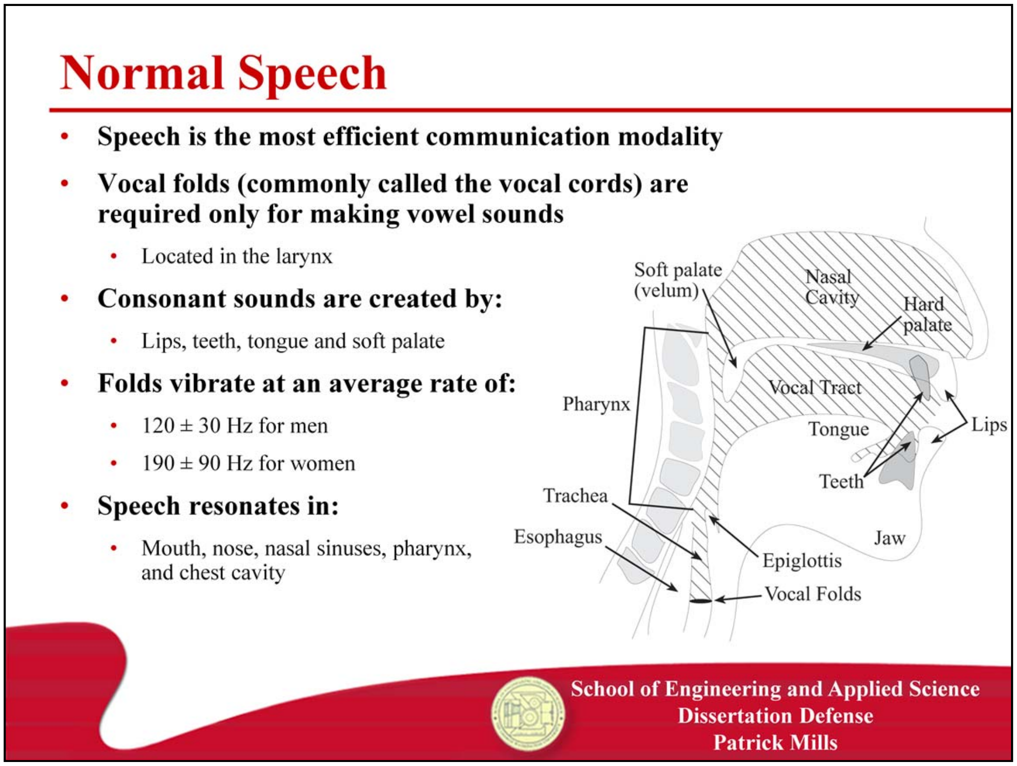
Audible ultrasound has been around for many years – originally developed by Pompei at MIT. His AudioSpotlight is a parametric array of transducers arranged in a small dish. Sound is modulated to ultrasonic frequencies; due to nonlinear interactions in the air and ear canal, the signal self-demodulates in the listener's ear. This gave me the idea for a new type of electrolarynx back in 2003; I originally proposed the idea as a project for Dr. Zara's medical instrumentation class. After a decade, I finally fully developed the idea into a working prototype.

Original Objectives

- **Show that a viable alternative to the classic electrolarynx exists**
  - No one else has suggested using ultrasonic difference waves as an alternative to a piston
  - Difference waves have been used in other areas (audio delivery, vibroacoustography)
- **Simulate a physical approximation of the airway to determine reasonable parameters for the physical device**
- **Build and test a physical device**
  - Used a custom distilled water phantom
  - Record frequency and SPL of generated fundamental frequency
- **Publish paper describing the concept**

These were the objectives I presented when we met in May 2013.

The main objective is to show that a viable alternative exists to the classic piston driven electrolarynx.



Speech is the most efficient communication modality; no other single modality comes close. In order to achieve greater efficiency, speech must be combined with other modalities.

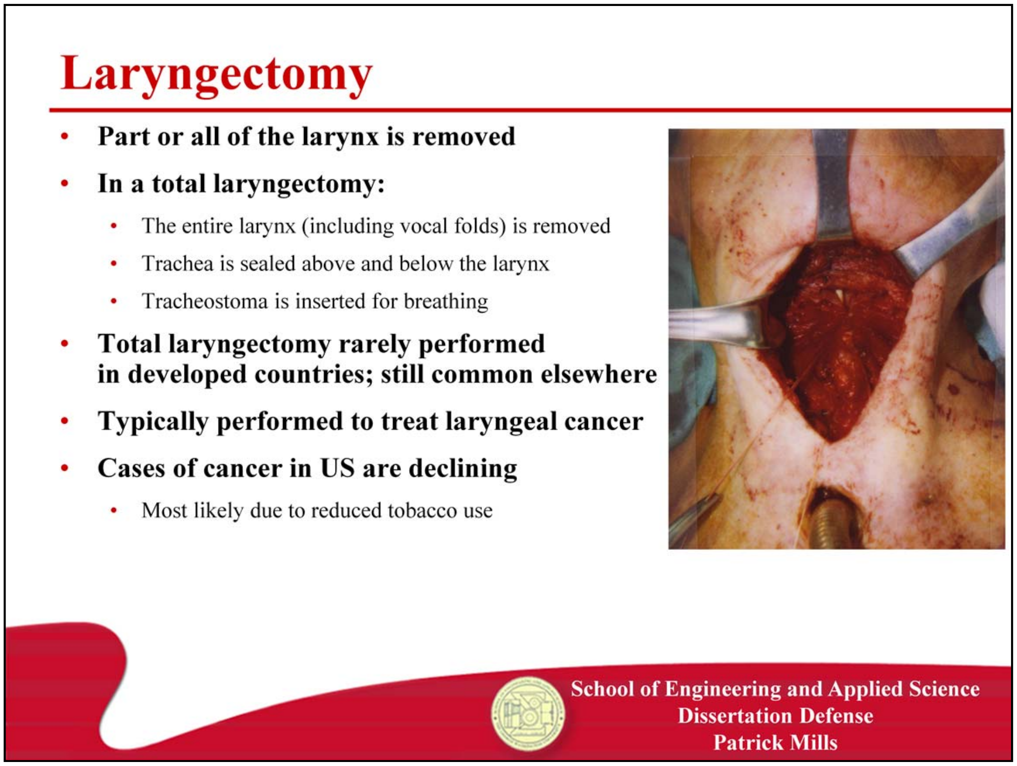
Loss of speech is a huge loss of self -> who we are is intimately connected to our "voice".

Vowels sounds are created by the vocal folds (commonly called the vocal cords) which are located in the larynx. The folds vibrate at about 120 Hz for men and 190 Hz for women.

Consonants are formed by shaping the voicing sounds left from vowels and plosives using the lips, teeth, tongue and soft palate.

Speech resonates in the mouth, nose, sinuses, airway and chest which is why you sound different when you have a cold.

All languages on earth require vowels; speech is not possible without them.



A laryngectomy is a complex operation in which part or all of the larynx is removed.

In a total laryngectomy:

The entire larynx (including vocal folds) is removed

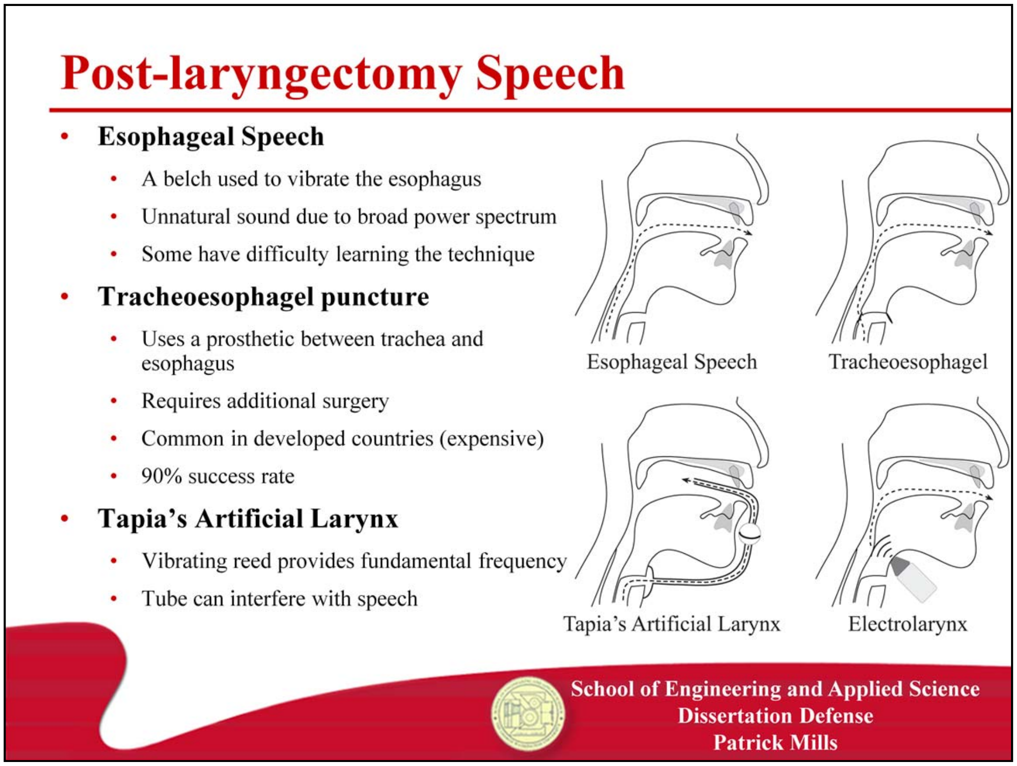
Trachea is sealed above and below the larynx

Tracheostoma is inserted for breathing

A total laryngectomy is rarely performed in developed countries (where the practice is to preserve as much tissue as possible – although recent studies have begun to question the efficacy of this practice); it is still common elsewhere in the world.

They are typically performed as treatment for laryngeal cancer; in the US cases of laryngeal cancer are declining (most likely due to reduced tobacco use), but are still rising in other countries (even developed ones such as Scotland).

Those treated often suffer post-surgical depression. Most are embarrassed about their post surgical voice. Many have issues at work and home due to their reduced ability to communicate effectively.



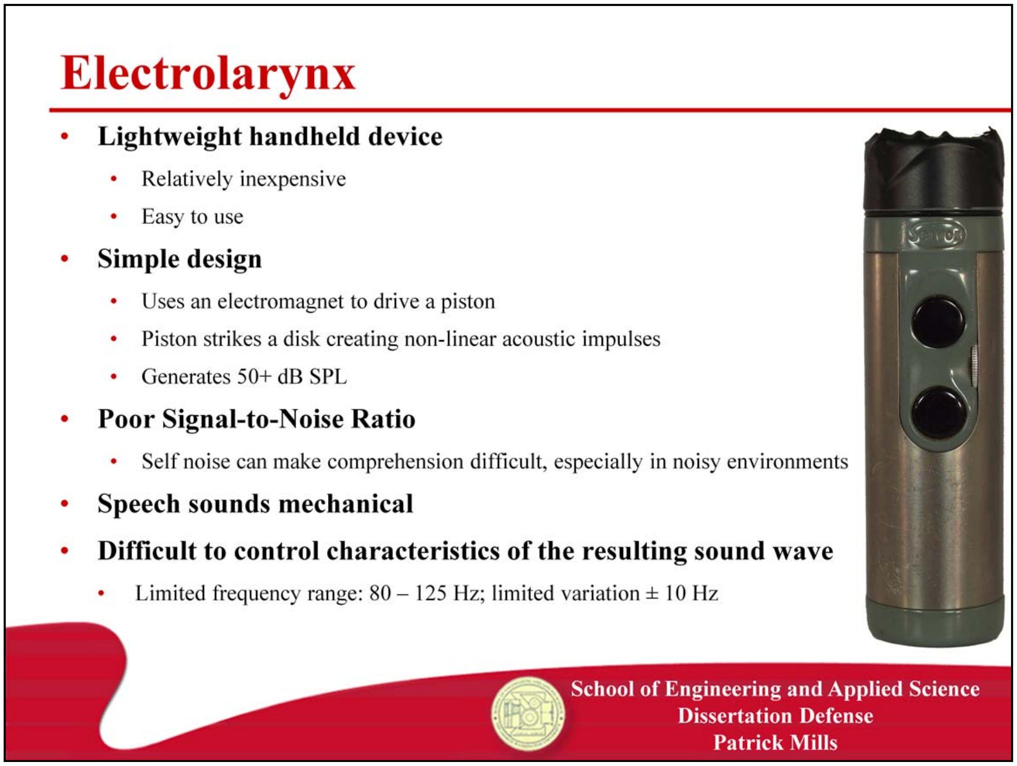
There are several techniques and technologies for restoring speech after surgery.

Esophageal speech is a biological only technique that uses a belch like action to vibrate the esophagus. Many people have difficulty learning the technique, but for those that can, the resulting speech is quite clear although unnatural sounding due to the broad power spectrum. Also, air is coming from the stomach, so speech length is limited to shorter bursts.

The Tracheoesophagel prosthetic is a surgically implanted one-way valve between the trachea and esophagus. In some devices, an embedded reed provides the voicing sound during a strong enough exhalation; without the reed, the PE-segment can be vibrated. The device is more common in developed countries due to the expense and requirement for additional surgery. When used, it has a 90% success rate. The prosthesis must be periodically removed and cleaned. There is also a risk of it detaching.

Tapia's is a type of pneumatic larynx; similar to the one first introduced in 1859. A funnel is attached to the tracheostoma; an attached tube contains a reed which vibrates to produce the voicing sound; the other end of the tube is inserted into the mouth. The tube can interfere with speech as it affects the movement of the lips, teeth and tongue.

The Tokyo Artificial Larynx is the last commercially produced pneumatic artificial larynx.



The electrolarynx is a lightweight handheld device that is relatively inexpensive (under about $1000) and is easy to use. Estimates indicate that 95% of all laryngectomees use an electrolarynx at some point in their recovery.

[Show the physical device] I have brought one with me today. This particular model is the popular Servox Inton.

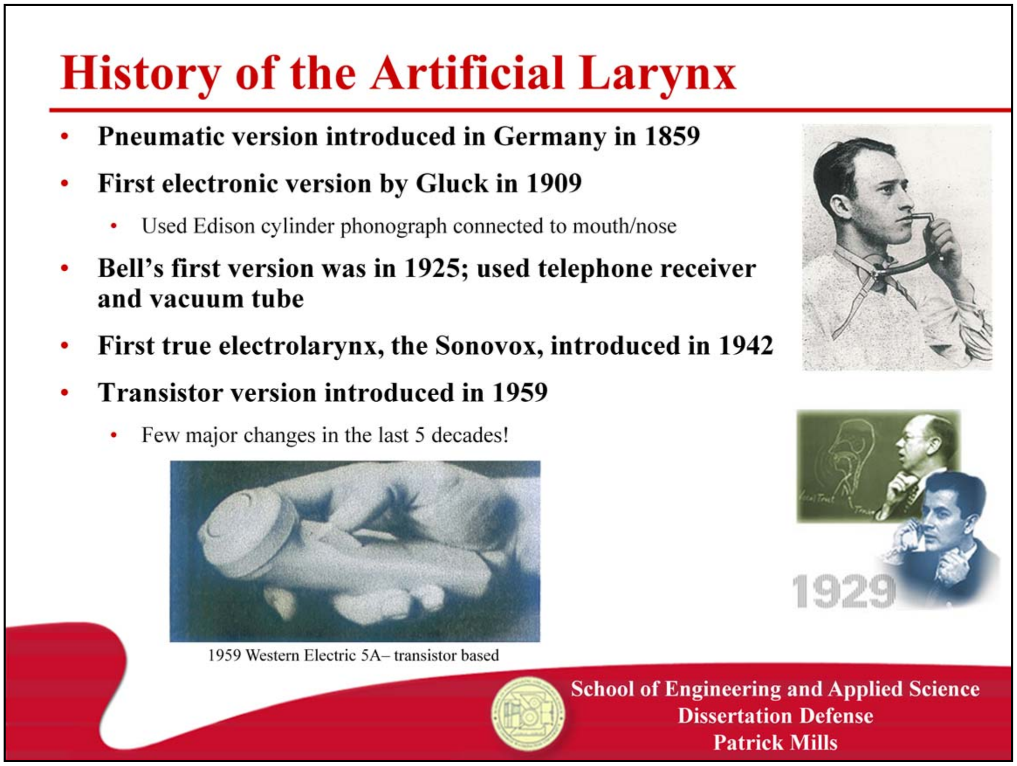
The design itself is quite simple consisting of a piston driven by an electromagnet which strikes a disc located at one end of the device. When pressed firmly to the neck, the coupler disc imparts acoustic impulses into the soft tissue where they mechanically couple with the airway to produce a voice sound. The majority of the device space is for the battery and piston mechanism.

The devices suffer from high self-noise: about 80 dB @ 30 cm when not coupled to the neck; 66 dB @ 30 cm when coupled

[Show the level of self-noise]

Modern versions do not have the coupler disc and piston physically connected in order to achieve a higher SPL; however, this results in a non-linear excitation that makes controlling the resulting waveform difficult. Therefore, most devices offer limited controls.

[Point out the two buttons for pitch selection; show the coupling slider]

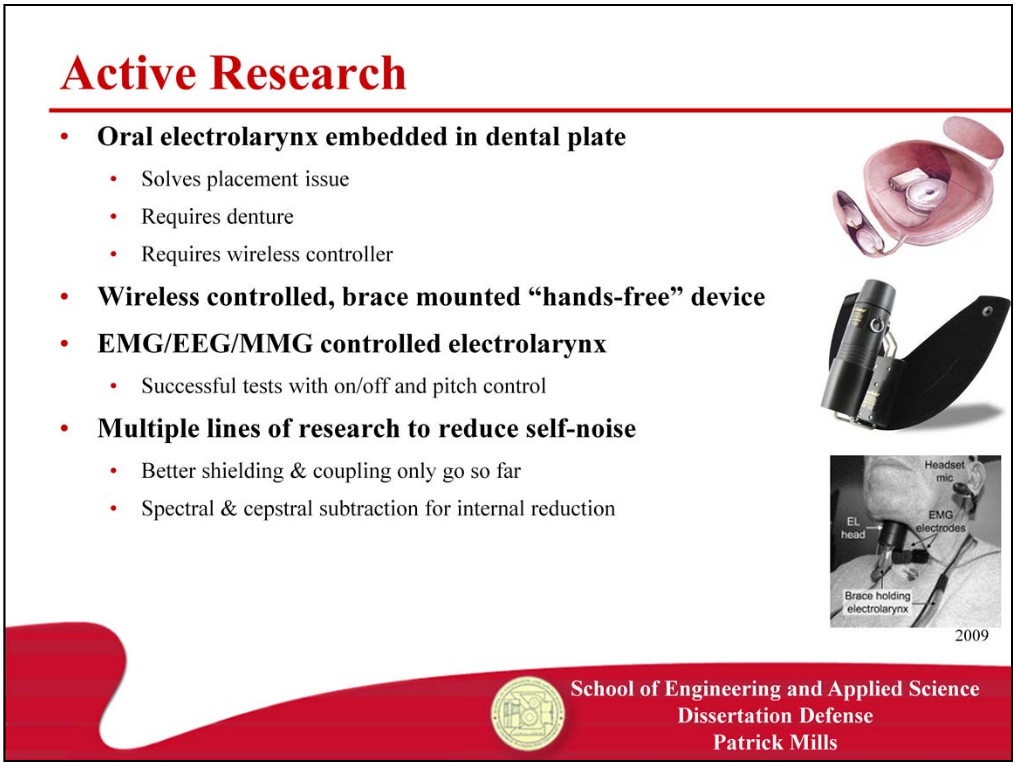


The first artificial larynx was a pneumatic version introduced in Germany in 1859. [see photo on the right]

The first electronic version by Gluck was introduced around 1909. It used an Edison cylinder phonograph with a recording of an opera singer loudly singing a vowel sound. The output was piped into the mouth or nose.

In 1925, the Bell company introduced their first version which was created when a friend of an executive underwent a laryngectomy and complained that he needed a better solution. The device used a vacuum tube where one plate was allowed to vibrate much like the vocal folds; the varying distance between the plates controlled the output to a telephone receiver. The device suffered from a very low SPL.

The modern electrolarynx was born in 1959 when the first transistorized version, the Western Electric 5A, was released. The device has changed little in the last 5 decades – the major advances are the non-linear coupling disc and better batteries.



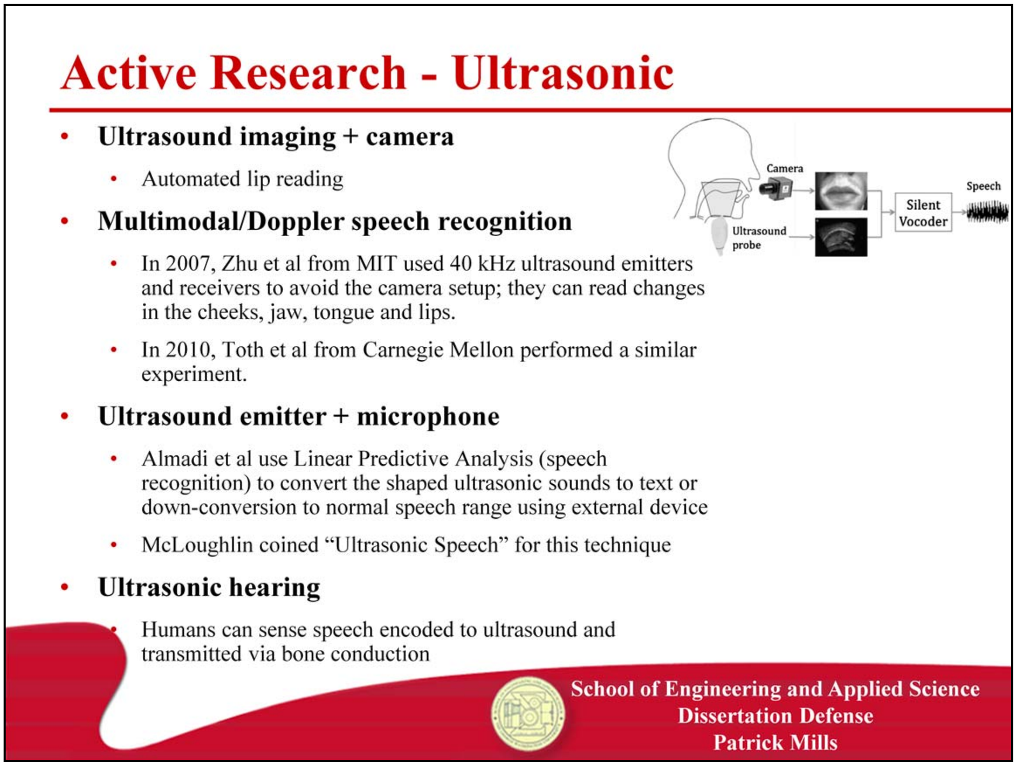
Given that most laryngectomees use an electrolarynx at some point in their recovery, and that worldwide the number of laryngectomies is still increasing, there is active research into improving the electrolarynx.

One solution is to embed the sound source in a dental plate; this solves the sometimes difficult placement issue, but requires a denture and still requires the use of a hand to operate the wireless controller.

A classic handheld electrolarynx mounted on a brace allows the user to position their neck on the device as needed; as with the dental device a hand is still necessary to operate the wireless controller.

The research I see as having the most potential as electronics continues its inevitable miniaturization and processors continue to increase in potential computing power is the use of the EMG to monitor the neck muscles, EEG to monitor brain activity, and/or MMG (mechanomyography) to monitor motion via accelerometers to determine when to activate the device. Research in this area has been ongoing since the 90s, but tests in 2012 showed the ability to not only turn the device on/off, handle pauses, but also limited pitch control. I am unaware of any commercial devices that utilize this technology at this time.

There are many lines of research into reducing the self-noise produced by the electrolarynx. Better shielding and coupling only go so far. Techniques adopted from speech recognition offer the potential to allow active noise reduction, spectral shaping and cepstral subtraction.



Research involving ultrasound takes many forms.

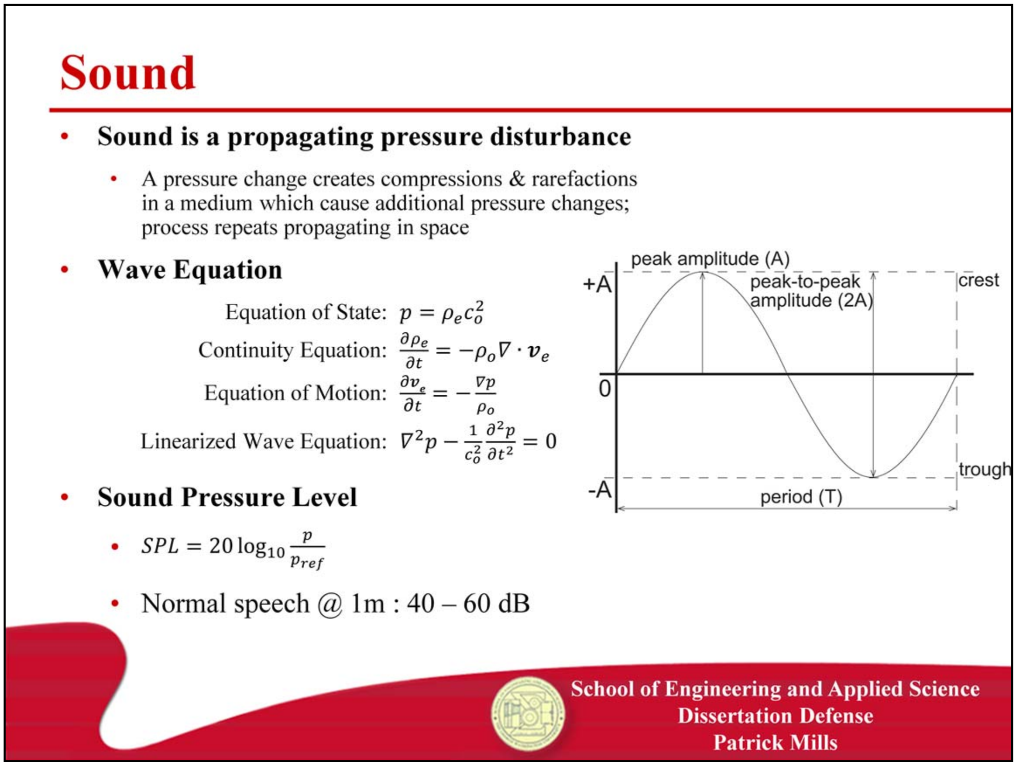
One solution uses ultrasound imaging with a camera to perform automated lip reading.

[see photo on the right]

Research at MIT and Carnegie Mellon replaced the camera with a 40 kHz ultrasound transducers; they are able to read changes in the cheeks, jaw, lips and even tongue using the Doppler effect. Combined with a microphone, they are able to perform automated lip reading.

Almadi et al pipe ultrasound directly into the mouth where it is shaped by normal processes. A microphone connected to an external device (like a smart phone) records the shaped ultrasound and converts it to text using speech recognition techniques or down-converts it to normal speech range. McLoughlin coined the term "Ultrasonic Speech" for this technique.

An interesting research line involves hearing restoration; when normal speech is shifted to ultrasonic frequencies and transmitted via bone conduction, humans can "hear" it – the technique offers the potential for partial hearing restoration for those with low frequency hearing loss.



Sound is a propagating pressure disturbance

A pressure change creates compressions & rarefactions in a medium which cause additional pressure changes; the process repeats propagating in space. Sound requires a medium to propagate.

The wave equation is an elegant expression of the propagation of wave pressure.

To derive the equation, we need three fundamental equations:

The Eqn of State where we assume pressure is a function of density alone

The Eqn of Continuity represents the conservation of mass : rate of mass out – rate of mass in + rate of buildup = 0

And the Eqn of Motion which represents the conservation of momentum for a fluid : acceleration = Force / mass

Combined they produce the wave equation, a differential equation describing the propagation of wave pressure (or velocity depending on derivation).

Looking at the photo on the right as reference, the fundamental properties of waves are:

Amplitude = a measure pressure generated by the wave

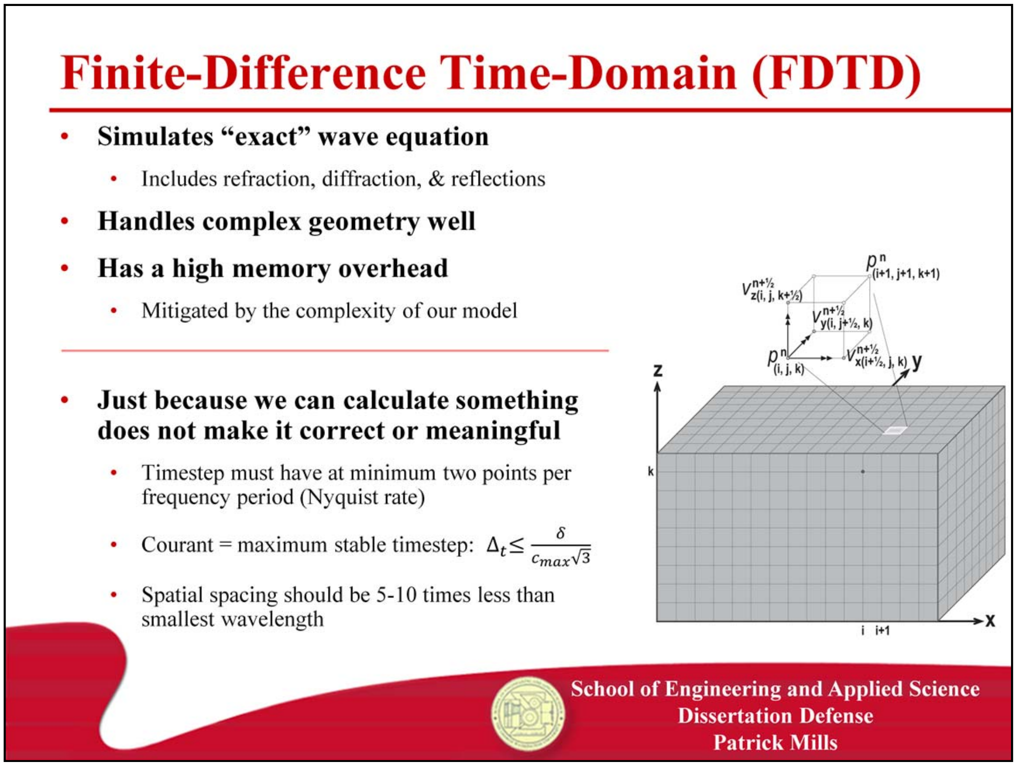
Crest & trough – the maximum and minimum amplitudes

The period is the time to complete one cycle and the frequency is the reciprocal of period.

The Wavelength is the distance between two phase related points within one period.

When comparing sounds, the SPL or Sound Pressure Level is typically used. It represents the sound pressure relative to a reference at a specific distance.

p_{ref}, is typically 20 µPa for air - representing the RMS sound pressure threshold of human hearing for a 1 kHz tone
for water 1 µPa is used



There are many methods for calculating the wave equation:

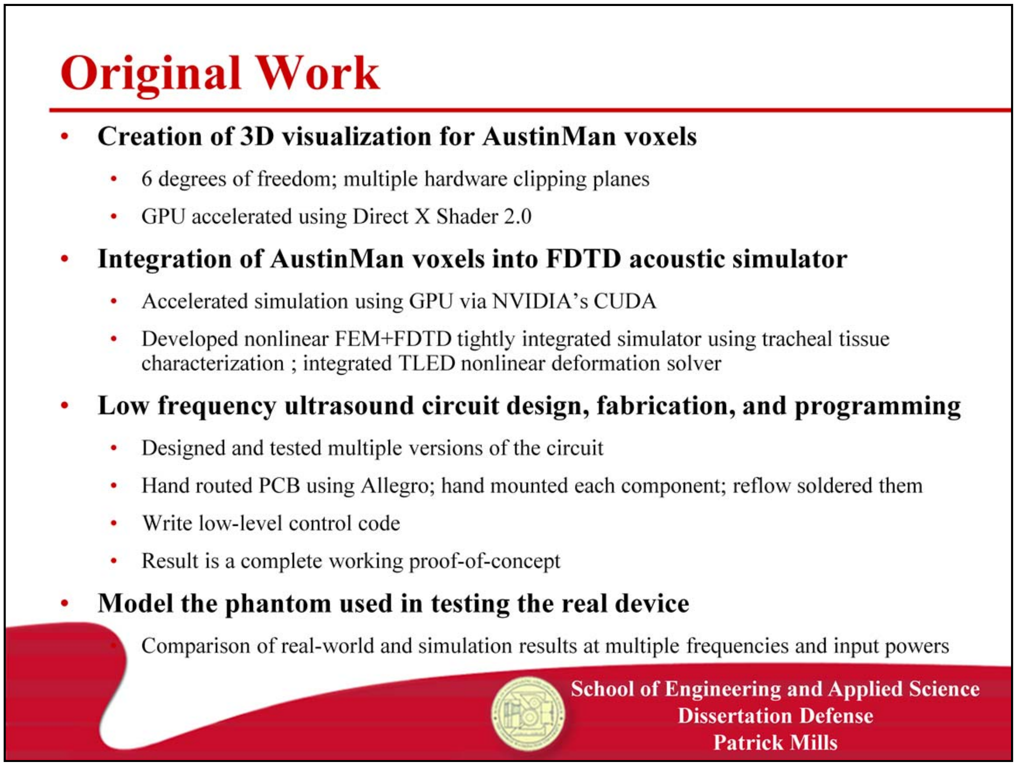
The Psuedospectral method offers high accuracy & low computational overhead but requires a relaxed grid that does not work well with complex fixed geometry.

Finite-element & finite-volume methods work with complex geometry, but have a complicated implementation.

FDTD works well with complex geometry but suffers from high memory and computational overhead, however, our model's memory requirements mask most of the memory issue; and the GPU mitigates the computational overhead as the FDTD is easy to massively parallelize.

With FDTD, the "trick" originally proposed by Yee, is to stagger the pressure and velocity fields in both space and time. Conceptually, we located the pressure nodes on the grid and the velocities on the axes. Start with initial pressures (scalar field); calculate velocities then leapfrog to pressures, etc. Each calculation is independent of the other in both time and space; this allows for great parallization.

[discuss calculation]



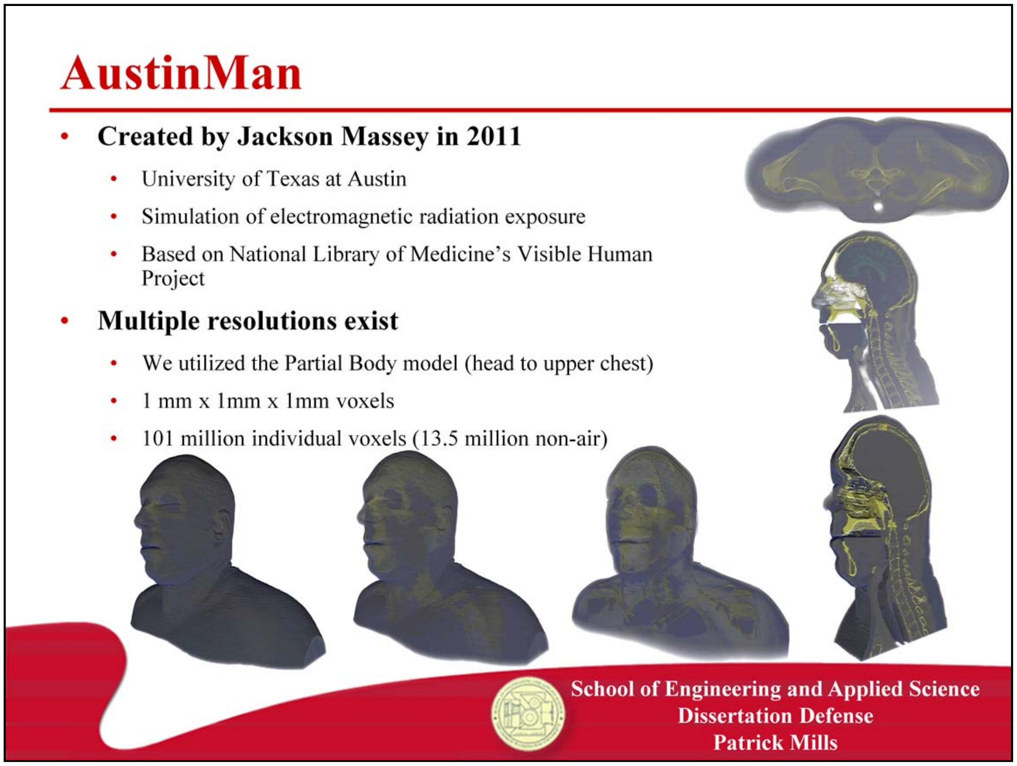
To complete my dissertation, I have:

Created a custom 3D renderer to visualize the AustinMan voxels. The software allows 6 degrees of freedom, multiple hardware clipping planes and is GPU accelerated using Direct X Shader 2.0 routines.

Created a custom acoustic FDTD simulator handling multiple materials and parameters and integrating the AustinMan voxels. The software is GPU accelerated using custom CUDA code. I extended this code to include a nonlinear Finite Element Model tightly coupled to the FDTD.

Designed and tested various circuits for hardware implementation. I then created a circuit diagram using Cadence Capture and created a PCB using Allegro. I had the PCB fabricated, placed the components and validated the circuitry. I wrote custom control code.

Using the hardware, I tested the device on a custom made phantom.



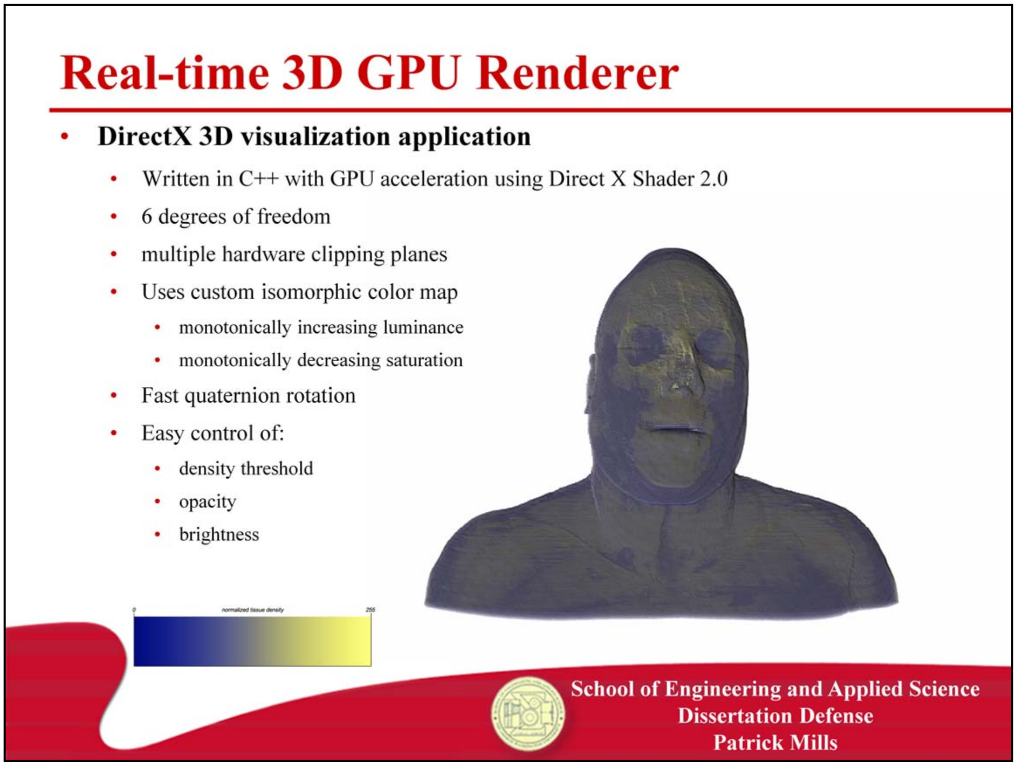
The AustinMan voxels were created by Jackson Massey as part of his thesis in 2011 which explored cellphone electromagnetic radiation exposure. A voxel is a volume pixel or 3D pixel.

He started with cryosection images from the National Library of Medicine's Visible Human Project. Using custom software and manual review, he classified tissue boundaries. After he graduated, he and a team at the University of Texas at Austin continued to refine the software and produce higher resolution data sets for both the Visible Man and Woman.

UT Austin shares those files with approved sources.

We utilized the v1.1 Partial Body model (which includes the head to upper chest) at the highest resolution ($1mm^3$).

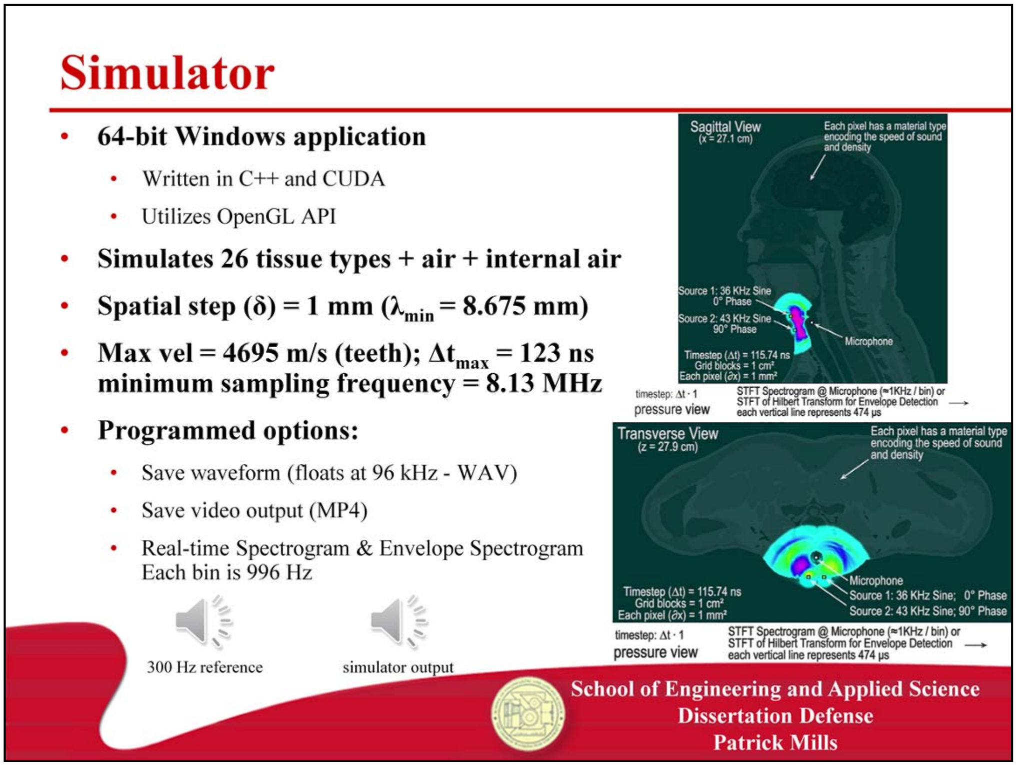
The data set consists of 101 million individual voxels (13.5 million non-air)



I implemented a custom extraction system to convert the UCD format voxels into a format my 3D renderer could utilize. The renderer supports 6-degrees of freedom and multiple hardware clipping planes and allows me to examine the data set.

[Show software]

I want to point out that a lot of work went into making this software usable given that we are displaying 101 million voxels in real-time; I have also done research on visualization and applied techniques to correctly color the body for optimal rapid comprehension. [discuss efficiency eg SolidWorks 2012 vs IntelliCAD 2001]



For the simulator, I started with a project by Ola Vikholt from Norway; he created a basic FDTD simulator as an undergraduate project, ported it to the GPU as his master's thesis, and continues to refine it in his job as an audio engineer. He has shared it with a limited number of researchers. His original model handles only air and a single material.

I ported the code to a 64-bit Windows application, fixed some tricky CUDA bugs, extended the system to support multiple tissue types, and upgraded the FDTD calculations to fourth order. The initial version operates on one slice of the model at a time.

Simulations were run on Intel i7 2.8 GHz with 8 cores and 24 GB of RAM

NVIDIA GeForce GTX 460 with 336 cores and 1 GB of RAM

The entire model requires 45 MB of space; however, the wave state requires 4 double precision variables taking a total of 1.4 GB.

In order to simulate using the entire model, we examined four possibilities:

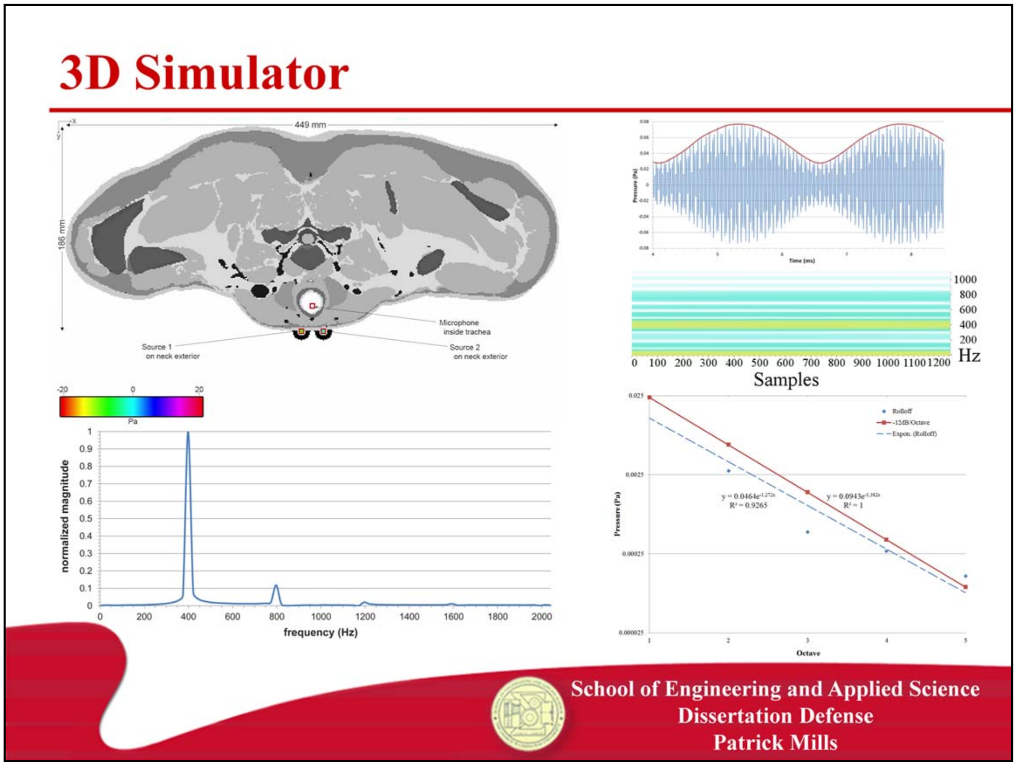
Switch from double to float precision – for FDTD only this doesn't really present a problem although it does for the nonlinear model (both from a precision & space perspective).

Use host mapped memory – store everything on the host and access it via the PCI bus. Too slow (a month of computation for one simulation)

segment the slices importing a segment from the host to device (plus a halo), calculating, then copying back to the host. (9 days – still too slow)

upgraded to NVIDIA GeForce GTX 660 Ti with 1344 cores and 3 GB of RAM – (17 hours)

We validated the results by comparing analytic calculations to the simulator results.



On average we simulate for almost 550K iterations (67 ms of real-time).
Steady state occurs after about 1500 iterations.

For 400 Hz, this gives us 26.8 cycles; (13.4 cycles at 200 Hz)

FFT uses 4096 samples @ 96 KHz.
For 400 Hz, this gives us 17 cycles; (8.5 cycles at 200 Hz)

3D Simulator Results

Beat Frequency	Frequencies (KHz)	SPL (dB)
400 Hz (179 mm)	40.0 & 39.6	61.47
200 Hz (172 mm)	40.1 & 39.9	62.37
200 Hz (179 mm)	40.1 & 39.9	46.94
200 Hz (183 mm)	40.1 & 39.9	58.86

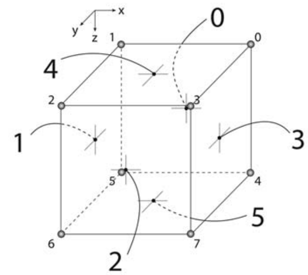
Example sound sources	SPL (dB)
Hearing threshold at 1 KHz	0
Quiet bedroom at night	20
Quiet library at 6’	30
Normal Conversation at 3’	40-60
Vacuum cleaner at 3’	70
Busy roadside at 32’	80
Jackhammer at 50’	95
Threshold of discomfort	120
Threshold of pain	130
Jet engine at 164’	140

Our results were not as high as we wanted, but very respectable for a first prototype device.

Comparing to sound level chart, our results are within the range of normal conversation at 3’.

Nonlinear Simulator

- **Used Holzapfel strain energy density function Tracheal Tissue Characterization**
  - Anisotropic viscoelastic characterization by
    Trabelis et al. & Malvè et al.
  - Implemented in CUDA
- **Used total Lagrangian explicit dynamic (TLED) FEM algorithm**
  - Apply loads and boundary conditions.
  - Calculate the deformation gradient.
  - Calculate the PK2 stress.
  - Calculate the linear strain-displacement matrix.
  - Calculate the element nodal forces.
    Acoustic radiation pressure · incident surface area
  - Calculate the displacement.
    Relate new volume to density



Beat Frequency	FDTD SPL (dB)	FEM SPL (dB)
400 Hz (179 mm)	61.47	58.40
200 Hz (172 mm)	62.37	62.28
200 Hz (179 mm)	46.94	46.87
200 Hz (183 mm)	58.86	58.91

School of Engineering and Applied Science
Dissertation Defense
Patrick Mills

We implemented a tightly coupled nonlinear version of our simulator.

We are interested in the reflection of incident waves occurring at the tissue / air boundary.

Since very little of the energy is transmitted across the boundary due to the impedance mismatch, a force is applied to the air facing boundary when the wave reflects.

We use the FDTD model to calculate the pressure & velocity up to the boundary; then calculate the force applied to the wall. This is then used to deform the tissue based on the anisotropic viscoelastic tissue characterization. A FEM is used to calculate this deformation.

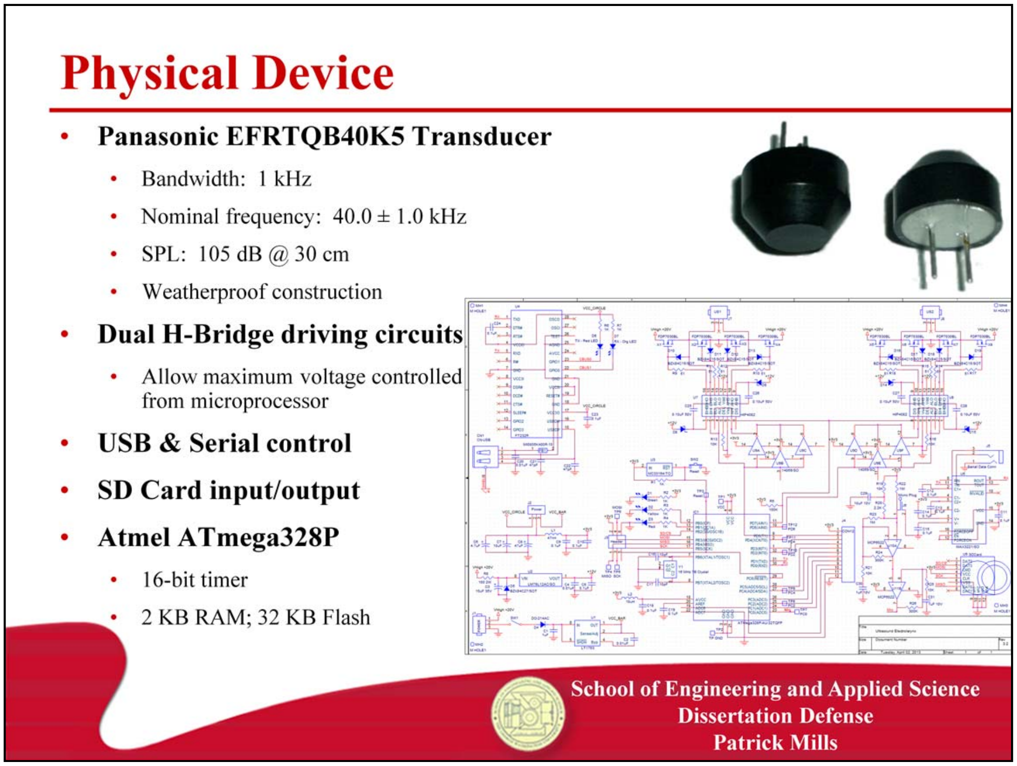
In our linearized model, we assumed that pressure is a function of density alone. We used this to couple the FEM back to the FDTD simulation. Basically tissue is incompressible; this is expressed in the Poisson ratio. A value of 0.5 is completely incompressible; tissue is often assume to be 0.49. If we look at the volumetric part of the tracheal characterization – a standard neo-Hookean model is used – we can then extract parameters to see what the Poisson ratio would be. This is a bit of a fudge as these parameters are only valid for homogeneous isotropic linear elastic materials; however, depending on our assumptions, we get a Poisson ratio of 0.485 to 0.4925 (0.49 fits nicely).

As the tissue is basically incompressible, its density will remain the same during deformation. However, air is highly compressible. And so we calculate the volume change of the computational cells adjacent to the tissue. We then update the air density and use that in the FDTD simulation.

As you can see in the output table, we did not see the mechanical resonance we were looking for; we expected to see higher output levels for the FDTD+FEM model than for the FDTD alone. So what went wrong? Short answer – nothing. Results were the same or a little lower than the 3D linearized simulations.

We implemented the worst-case solution. The FEM elements are disconnected; this localizes any deformations. We also only consider the boundary tracheal muscle layer – non-boundary points are fixed; deeper layers and tracheal cartilage are ignored. The best-case would be a connected grid, with flexing multi-layer deep. However, we had no documentation/evidence of this – there are no studies of this type of modelling that we could find – the studies we did find focus on modelling airflow within the vocal tract; this is where the tissue characterization came from. We also implemented the strong coupling indirectly (as a density change); we could have modelled this as an active pressure wave – again, we have no evidence to support this implementation – even just watching the vocal folds is difficult, much less the tracheal tissue.

Furthermore, our model has limited resolution; a finer resolution would include cilia, epithelial cells, mucosa, fibrous membrane, etc. However, we have no characterization of those. And the computational resources required to do the model would be large – funding would be required to enable such a model. The existing model took 7 months of research and implementation, but runs on a few thousand dollars worth of hardware.



As the simulation showed encouraging results (we actually see a signal within the vocal tract – it is not completely blocked due to the impedance mismatch), we moved forward with creating an actual device.

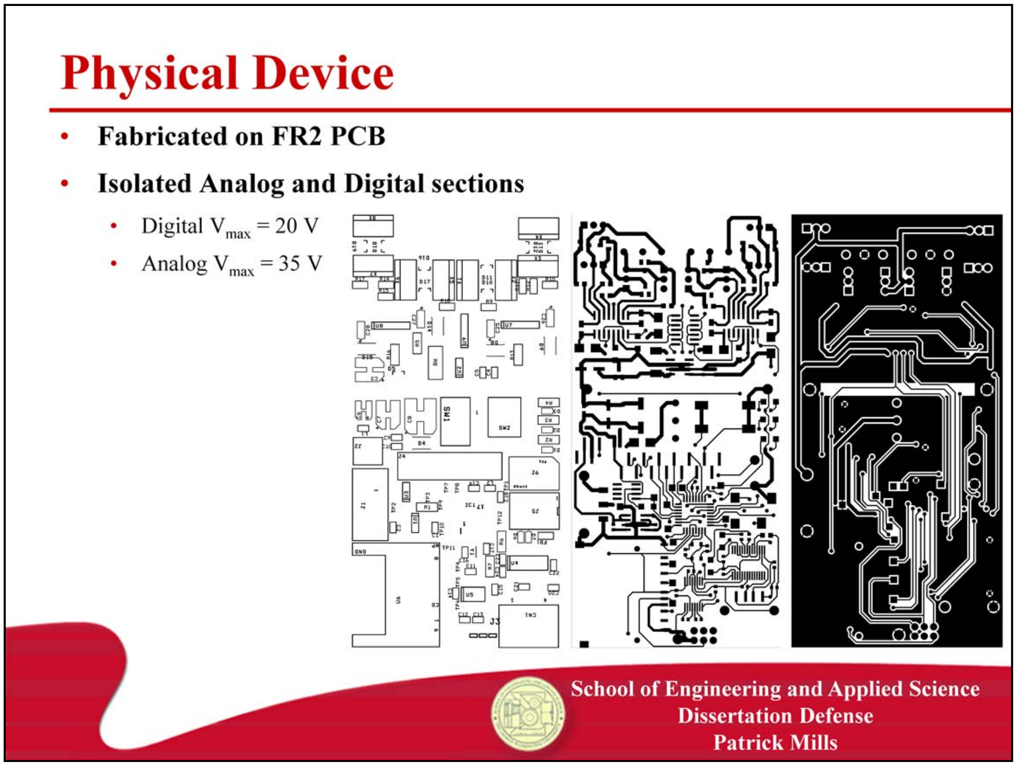
I identified suitable transducers, designed analog driving circuitry around their specifications.

Integrated digital control circuitry including multiple communication (USB/serial) and debugging options (JTAG).

Looking forward to a commercial device, I used my experience with robotics to implement a single power solution.

An H-bridge allows flipping a single supply voltage between the output terminals of the transducer.

Designed circuit layout in Cadence Capture



Circuit layout brought into Allegro for layout.

Created custom part packages where required (MicroSD, USB, jumpers, terminal block, etc)

Manually routed traces observing the need to isolate digital, analog and power.

Ground break isolates analog and digital

used wide power/ground traces, eliminated ground loops, minimized inductance

placed bypass capacitors as close to components as possible

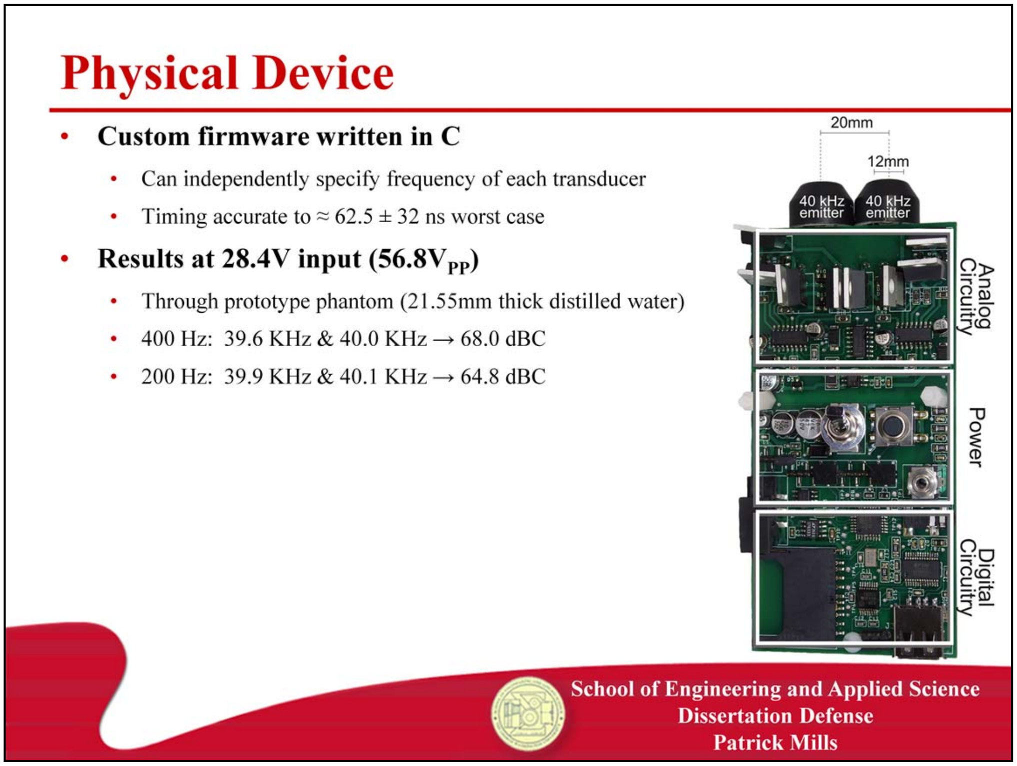
Testing: unconnected pins, bridges, solder balls / splatters, tombstones, grainy/cold joints

tested traces & verified parts using multi-meter & scanning tweezers

Powered digital first; set processor options for external clock using JTAG port & ICE programmer

Tested digital outputs using Saleae Logic16

Tested analog outputs using HP 1741A oscilloscope



We wrote the code in C and analyzed the assembly output to calculate the require cycles for critical interrupt service routines.

We tested 7 algorithms, using a combination of software & hardware features.

Ultimately, we used the Output Compare toggle to automatically trigger the wave transition (high to low / low to high) based on an 8-bit clock.

We used an Interrupt Service Routine to reset the clock timer. ISR takes only 19 cycles!

Clock is 16 MHz in prototype device (could be upgraded to 20 MHz without design change).

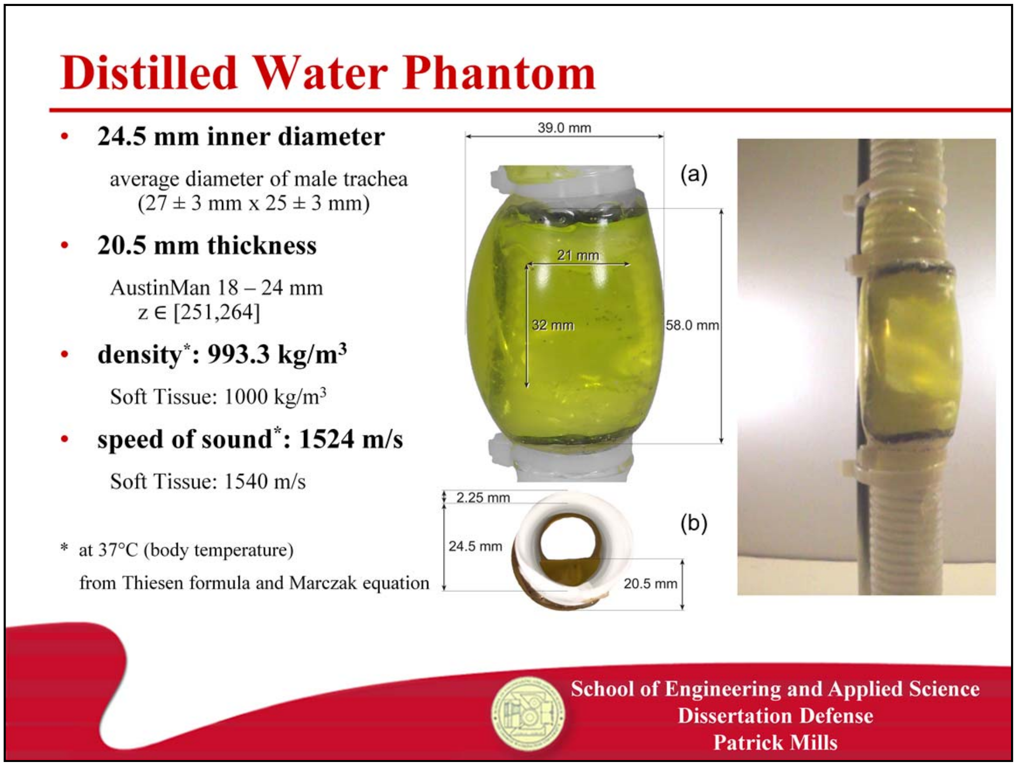
External clock is high-accuracy with a total error (tolerance + stability) of ± 10 ppm ± 10 ppm = ± 20 ppm or (20 µs @ 16 MHz)

Our output is accurate to within 62.5 ns (half clock cycle per edge worst case) as long as the ISR has time to complete.

If we assume worst case, we need 2 * 19 = 38 cycles; and with 8 bits for each half wave we can handle up to 512 cycles. So, our output signals can be between 31 KHz and 421 KHz.

In the area of our interest (40 ± 0.5 KHz), we get error digitizing our timing counter.

This leads to a worst case error of 31.25 ns (0.13%).



To test our device, we designed a distilled water phantom.

We used:

latex torus (child's toy) and filled it with distilled water

mounted that to a piece of split loom polyethylene tubing with a hole cut out.

The vocal tract is basically a sound tube, and picking an appropriate diameter and length gets us a very good approximation.

(we miss some of the finer aspects as we do not have the mouth & nasal cavities, but these are not in our focus area at this time anyway.)

We have good agreement between the model and soft tissue for density & speed of sound; attenuation of distilled water is not as great as soft tissue.

Device Results

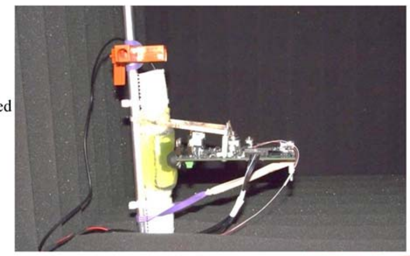
Voltage (V)	Vpp	dBA	dBC
12.5	25	52.5	54.0
15	30	58.4	59.2
20	40	63.2	63.9
25	50	66.0	66.8
26.3	52.6	66.2	67.3
28.4	56.8	66.4	67.6
30	60	66.8	68.0

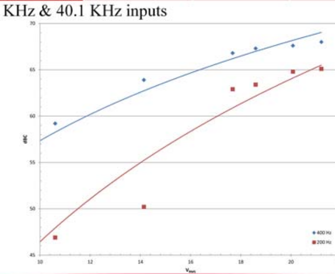
39.6 KHz & 40.0 KHz inputs

Voltage (V)	Vpp	dBA	dBC
12.5	25	34.4	45.8
15	30	36.2	46.9
20	40	46.2	50.2
25	50	62.5	62.9
26.3	52.6	62.8	63.4
28.4	56.8	63.7	64.8
30	60	63.8	65.1

39.9 KHz & 40.1 KHz inputs

Anechoic sound Chamber with distilled water phantom





School of Engineering and Applied Science
Dissertation Defense
Patrick Mills

We also built a 2’ x 2’ x 2’ sound chamber using a corrugated box with alternating 2” acoustic tiles inside and 4 mm acoustic foam on the exterior. The box was raised above the solid table using foam to further isolate low-frequency coupling.

All measurements were performed within the chamber. Even so, limiting low frequency noise is extremely difficult. We turned off lights, shutdown HVAC, turned off electronics with fans. We were able to detect a truck idling two blocks away. Ultimately, we performed measurements during times when traffic was light.

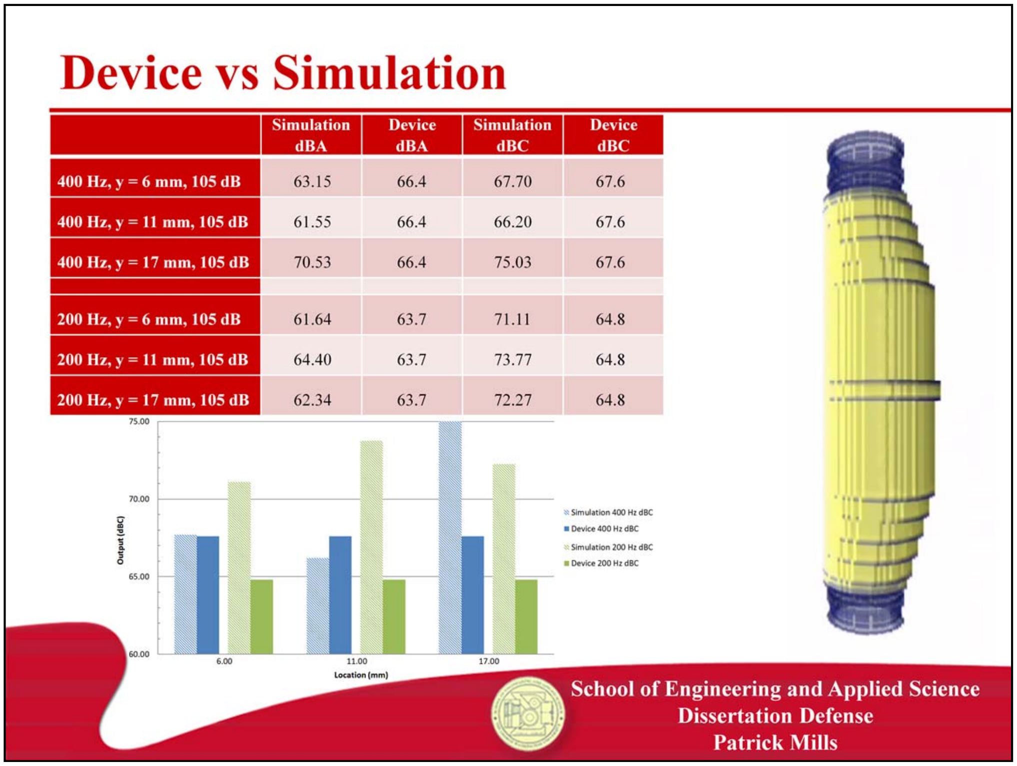
Measurements were taken using a Sper Scientific 850014 sound meter in slow mode (1 s sample time @ 16 KHz).

Coleman *et al.* : minimum SPL at 15.25 cm for F_0 = 48 dB (max = 126 dB)

Schindler *et al.* : tracheoesophageal voicing at 30 cm - minimum of 50 ± 4.8 dB and a maximum of 68 ± 4.7 dB.

The vocal tract is a filter which spectrally shapes the fundamental frequency into the voice, amplifying frequencies near resonances and dampening others.

However, even if we assume the worst case and apply the inverse square law where the energy dissipates uniformly in inverse proportion to the distance, our observed levels are still within the error bounds of both studies.



As we cannot be sure where the microphone is located in the phantom, we performed the simulation with three microphone locations:

front (y = 17mm) ; middle (y = 11 mm) ; back (y = 6 mm)

Best match is back…

We are not sure why the 200 Hz simulation and real-world tests differ so much more than 400 Hz.

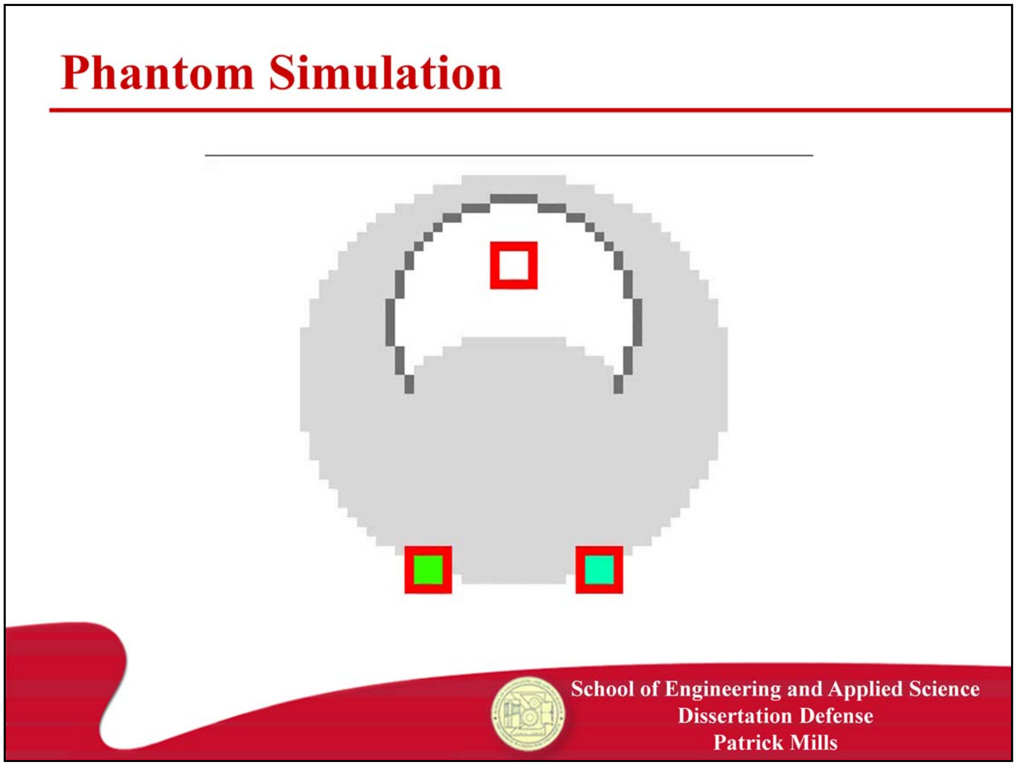
It could be:

an interaction with the torus latex coating (latex is not simulated)

A timing issue – we used a half cycle timing technique to allow 50 Hz resolution that results in longer off than on. Rewriting the routine to have more on than off might help.

An issue with our sound meter at lower frequencies

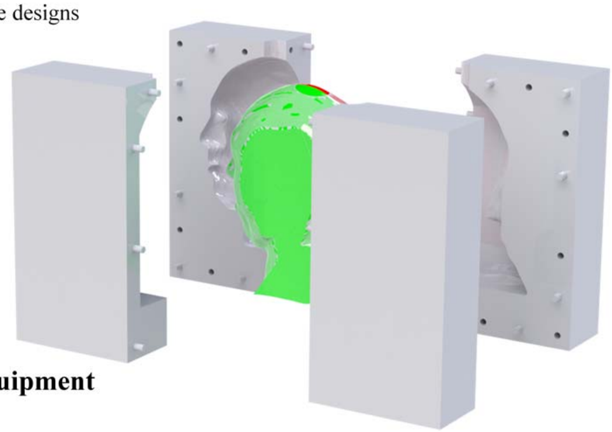
???



Here we have a short video of the 3D simulator running on the phantom model.

Future Work

- **More accurate timing**
  - Use 20 MHz chip – no design changes required
  - Use faster processor in future designs
- **Alternative transducers**
  - Ultimately custom MEMS
- **AustinMan Phantom**
- **Transient only output**
  - Vary output to avoid steady state – higher output
- **Revisit FEM model**
  - Implement connected mesh
- **Use better monitoring equipment**

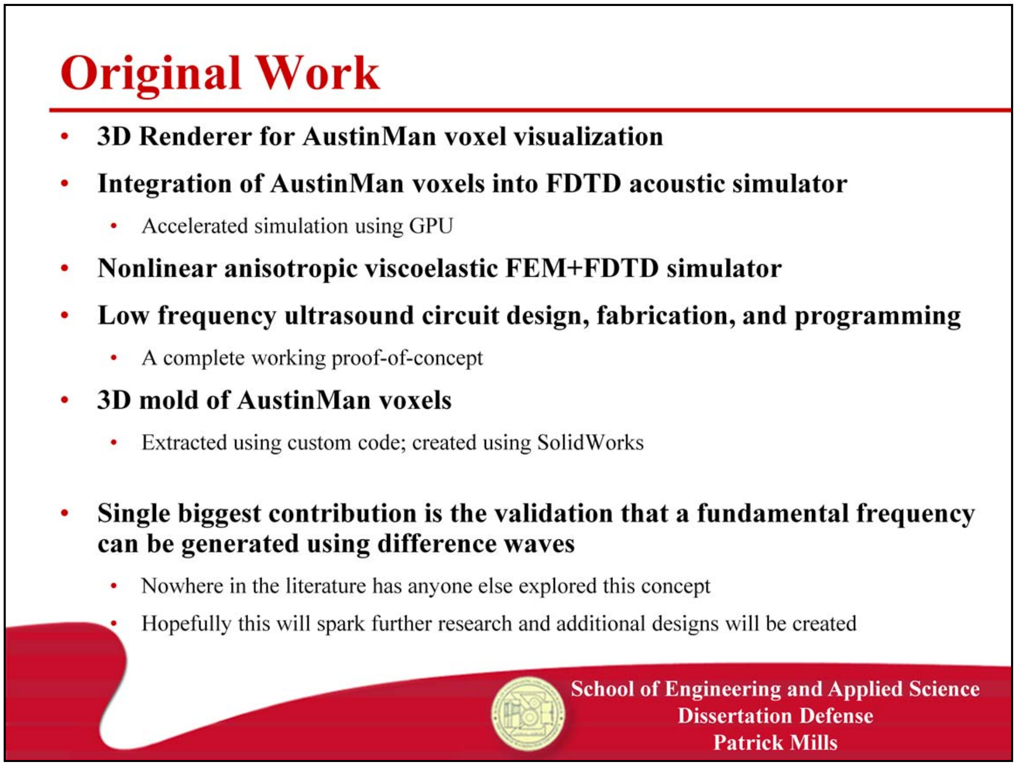


School of Engineering and Applied Science
Dissertation Defense
Patrick Mills

Future Work

- **EMG activation**
- **Inductive charging**
- **Single location optimized parametric transducer**
- **Small size, light weight**

- **My greatest hope is that the work so far inspires the creation of a medical device that can be FDA approved safe & effective**
  - Light-weight, comfortable, hands-free operation
  - Affordable
  - Better control of fundamental frequency parameters
  - Ultimately helping people live happier more fulfilled lives



In working on this dissertation, I have invested thousands of hours to explore an idea that has not been documented anywhere else.

After conceiving of the idea, I researched and when nothing existed, I developed a plan to see if the idea was viable.

I developed a simulator which showed that the concept should work.

I then developed custom hardware and validated that the idea does indeed work in the real-world.

Words mean more than what is set down on paper. It takes the human voice to infuse them with deeper meaning.
Maya Angelou

Questions?

School of Engineering and Applied Science
Dissertation Defense
Patrick Mills

The human voice is the organ of the soul.
Henry Wadsworth Longfellow

Thank You

Patrick Mills
patmills@gwmail.gwu.edu